Cancer Epidemiology 41 (2016) 139–151

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.canceepidemiology.net

Resolving uncertainty in the spatial relationships between passive benzene exposure and risk of non-Hodgkin lymphoma

Jeffrey M. Switchenko[a,*], Catherine Bulka[b], Kevin Ward[c,d], Jean L. Koff[b], A. Rana Bayakly[e], P. Barry Ryan[f], Lance A. Waller[a], Christopher R. Flowers[b]

[a] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[b] Department of Hematology and Oncology, Winship Cancer Institute, Emory University, Atlanta, GA, USA
[c] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[d] Georgia Center for Cancer Statistics, Atlanta, GA, USA
[e] Georgia Department of Public Health, Atlanta, GA, USA
[f] Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA

ARTICLE INFO

Article history:
Received 18 September 2015
Received in revised form 5 January 2016
Accepted 7 January 2016
Available online 2 March 2016

Keywords:
Spatial epidemiology
Lymphoma
Environmental epidemiology
Toxic release sites
EPA

ABSTRACT

Background: Benzene is a known occupational carcinogen associated with increased risk of hematologic cancers, but the relationships between quantity of passive benzene exposure through residential proximity to toxic release sites, duration of exposure, lag time from exposure to cancer development, and lymphoma risk remain unclear.

Methods: We collected release data through the Environmental Protection Agency's Toxics Release Inventory (TRI) from 1989 to 2003, which included location of benzene release sites, years when release occurred, and amount of release. We also collected data on incident cases of non-Hodgkin lymphoma (NHL) from the Georgia Comprehensive Cancer Registry (GCCR) for the years 1999–2008. We constructed distance-decay surrogate exposure metrics and Poisson and negative binomial regression models of NHL incidence to quantify associations between passive exposure to benzene and NHL risk and examined the impact of amount, duration of exposure, and lag time on cancer development. Akaike's information criteria (AIC) were used to determine the scaling factors for benzene dispersion and exposure periods that best predicted NHL risk.

Results: Using a range of scaling factors and exposure periods, we found that increased levels of passive benzene exposure were associated with higher risk of NHL. The best fitting model, with a scaling factor of 4 kilometers (km) and exposure period of 1989–1993, showed that higher exposure levels were associated with increased NHL risk (Level 4 (1.1–160 kilograms (kg)) vs. Level 1: risk ratio 1.56 [1.44–1.68], Level 5 (>160 kg) vs. Level 1: 1.60 [1.48–1.74]).

Conclusions: Higher levels of passive benzene exposure are associated with increased NHL risk across various lag periods. Additional epidemiological studies are needed to refine these models and better quantify the expected total passive benzene exposure in areas surrounding release sites.

1. Introduction

Benzene, toluene, ethylbenzene, and xylene (BTEX) are volatile organic compounds (VOC) that are typically found in petroleum products, coal tar and various chemical product formulations, and have been associated with increased cancer risk [1]. Among these, benzene is a VOC that has been consistently linked to hematologic cancers such as leukemia and lymphoma through occupational exposure [2–6]. Benzene is a widely used chemical that ranks in the top 20 chemicals for production volume and is used in the production of some types of rubbers, dyes, pesticides, lubricants, and detergents [7]. Studies have shown that subjects in occupations exposed to low levels of airborne benzene exhibit increased incidence of DNA methylation alterations common in acute myelogenous leukemia and other cancer tissues [8], in addition to lower levels of white blood cell and platelet counts [9].

The US Environmental Protection Agency (EPA) estimates that the main routes of benzene exposure occur through the air, via cigarette smoking and exposures from consumer products, car emissions, traffic exhaust fumes, and gas stations [10]. In addition,

* Corresponding author at: 1518 Clifton Road NE, Atlanta, GA 30322, USA.
E-mail address: jswitch@emory.edu (J.M. Switchenko).

http://dx.doi.org/10.1016/j.canep.2016.01.008

air around some toxic release sites may contain higher levels of benzene than other areas and thus contribute to the amount of passive benzene to which individuals are exposed [7].

The relationship between passive benzene exposure and hematologic cancers is less certain than for occupational exposure. Among hematologic malignancies, NHL is the most common. In 2015, an estimated 71,850 people in the US will be diagnosed with NHL, and 19,790 will die from this cancer [11]. For reasons that remain unclear, NHL incidence rates increased over the last half of the 20th century and only recently stabilized. Although the descriptive epidemiology of NHL has been well characterized using population-based cancer registry data over the last several decades, the etiology of NHL and its specific subtypes is less well understood [12]. To address this problem, InterLymph engaged in a worldwide project to pool case-control studies and perform pooled analyses to maximize the statistical power for identifying risk factors across NHL subtypes. A recent series of publications identified environmental, lifestyle, and clinical risk factors for several NHL subtypes [13–16] and recent genome-wide association studies (GWAS) identified single nucleotide variants associated with increased risk of diffuse large B cell lymphoma [17–23], the most common NHL subtype. Despite these recent seminal advances, relatively little is known about the spatial epidemiology of NHL. Although some studies support links between toxic exposures and NHL incidence, others do not, and thus considerable controversy remains [24–26]. The series of InterLymph studies previously mentioned also identified etiologic commonality across NHL subtypes and highlighted occupational history as linked to NHL [15].

To improve our understanding of the relationship between lymphoma risk and passive exposure through proximity to release sites, we previously collected data from the EPA's TRI and modeled the number of lymphoma cases as a function of indirect exposure to benzene using mean distance to benzene release sites in the state of Georgia [27]. This research identified passive benzene exposure as being associated with increased risk of NHL, but failed to clarify its impact on NHL risk in terms of quantity of exposure, lag time from exposure to cancer development, and duration of exposure.

Our prior model simplified estimation of residential exposure patterns by determining the average distance from all benzene releasing sites for a given location. While this approach identified associations between benzene exposure and increased lymphoma risk, mean distance from release sites remains a crude measure that does not take into account the magnitude of passive exposure. An individual's personal exposure to VOCs is related to indoor and outdoor sources, including points of release such as TRI facilities and non-point releases such as on-road, secondary, and background. Although the contribution of point sources to the outdoor concentration of a VOC and to an individual's total exposure may be small, differences in VOC release amount and proximity may result in distinct levels of risk for populations with varying degrees of exposure. We sought to examine the collective impact on the relationships between residential benzene exposure and NHL risk as influenced by distance from TRI release sites, amount of benzene released per site, and lag time from the period of release.

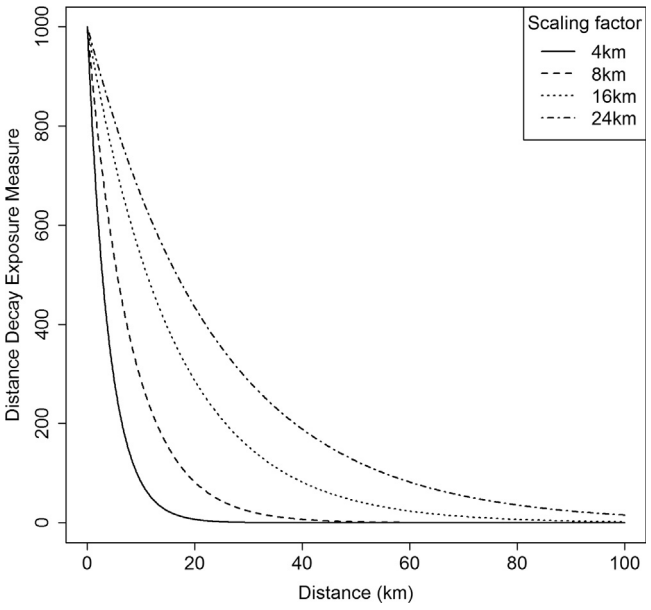

Fig. 1. Effect of scaling factor on the measure of distance decay exposure over increasing distances. Exposure at the point source (Distance = 0 km) is assumed to be 1000.

2. Data

We collected lymphoma incidence data from the GCCR for patients diagnosed with NHL from 1999 to 2008, benzene release data within Georgia from the EPA's TRI from 1989 to 2003, and state population characteristics from United States Census Bureau data for the year 2000. In the 2000 US census, there were 1618 census tracts within Georgia, of which 1616 had available population and demographic data. Data on sex, age, and race were obtained from Summary File 1 from the Census 2000 Data for the United States [28]. Georgia tract boundaries obtained from the Census Bureau's 2000 TIGER/Line files [29] were utilized for the purposes of allocating GCCR cases to Georgia census tracts. Census data for median year moved into residence (MYMI) were collected in our previous study, but were not found to be associated with NHL risk [27]. As a result, MYMI was not considered for further modeling purposes. Additionally, we collected Summary Files 3 and 4 socioeconomic status (SES) Sample Data, specifically the census tract level estimates for percent of the population older than 25 who are high school graduates and median income, in order to determine whether adjustments for these characteristics altered our findings [28]. All data were aggregated to the census tract level. Data collection was approved by the Emory University Institutional Review Board, the Winship Cancer Institute Clinical and Translational Review Committee, and the Georgia Department of Public Health Institutional Review Board.

2.1. Georgia Comprehensive Cancer Registry Data

From 1999 to 2008, the GCCR identified 12,716 NHL cases among adults $\geq$20 years of age living in Georgia at the time of diagnosis, of which 11,355 were successfully geocoded. Gender, race, and age-specific national NHL rates were obtained using data from SEER*Stat Version 7.05 [30]. Based on the demographic structure of each tract, we estimated the expected number of cases for each tract using these incidence rates. Thirty-two cases (0.28%) without gender, race, or age were excluded from further analysis. Lymphoma subgroups and subtypes were defined using ICD-O-3 codes based on the proposed World Health Organization-based nested classification of malignant lymphoid neoplasms for epidemiologic research from the International Lymphoma Epidemiology Consortium (InterLymph) [31].

2.2. Toxics Release Inventory Data

Facilities are required by the EPA to report their releases for certain toxic chemicals if they meet thresholds defined by Section 313 of the Emergency Planning and Community Right-to-Know Act: namely, if a facility is in a specific industry sector, employs 10 or more full-time equivalent employees, and manufactures or processes more than 25,000 pounds (11,500 kg) of a TRI-listed chemical or otherwise uses more than 10,000 pounds (4500 kg) of a listed chemical in a given year [32]. Release information included geographic coordinates of the release site, amount of on-site disposal or other release, mode of release (e.g., total air emissions, surface water discharges, etc.), and year of release [33]. The output "Total On-site Disposal or Other Releases" was used as the amount of on-site release for each site. For benzene, the amount of surface water discharge across all sites was negligible. From 1989 to 2003, 22 facilities in Georgia reported some disposal or release of benzene, including 3 facilities reporting benzene releases for all 15 years. The total amount of benzene released for each facility ranged from 10 kg to 1.48 million kg over this time frame.

3. Statistical methods

3.1. Exposure estimation—distance decay

We assumed that amount of benzene exposure was inversely proportional to the distance from benzene release point sources. We utilized an exponential decay function commonly used for estimating decreasing amount of exposure with increasing distance from a point source, in which the rate of decay is mitigated by a scaling factor that controls how gradually or quickly the decay occurs. With this approach, the contribution of a site to the estimated total passive benzene exposure in a census tract decreases with increasing distance from that site. We also assumed that the amount of exposure for a census tract due to benzene release from a toxic release site, independent of other sites, was related to the total amount of release from that site during the time period examined. The total amount of exposure for a census tract was defined as the cumulative exposure for that region from all sites with releases during that period.

Thus, we defined the exposure decay function as:

$$x_i = \sum_j R_j \exp\left(\frac{-d_{ij}}{b}\right)$$

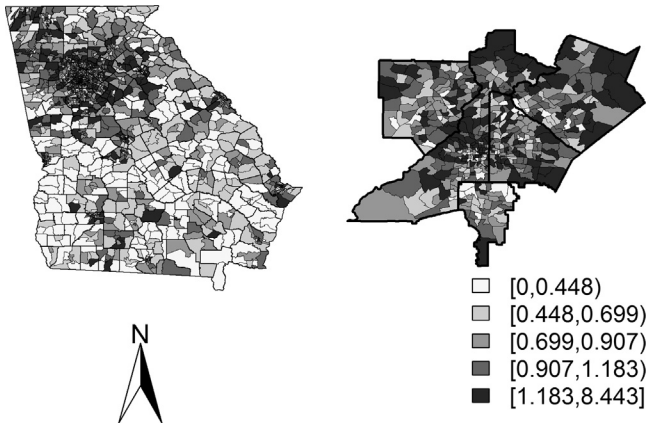

Fig. 2. Standardized incidence ratios for non-Hodgkin lymphoma, Georgia (left) and metro Atlanta (right).

where x_i is the cumulative amount of exposure for tract i,R_j is the amount of toxic release at release site j, d_{ij} is the distance between the centroid of tract i and location of release site j, and b is the scaling factor. Distance d_{ij} was calculated based on the haversine formula [34] for measuring great-circle distances from latitudinal and longitudinal coordinates. Thus, x_i represents the total exposure for a tract from all contributing release sites in the state as a function of distance from the release site and amount of release from the site during the period under consideration.

Exposure was then categorized into a discrete variable for analysis. A 5-level exposure variable was created using quintiles, with 5 equal-sized data subsets. Scaling factors of 4 km, 8 km, 16 km, and 24 km were explored in order to determine whether the chosen scaling factor influenced the relationship between exposure and disease risk. The scaling factor describes a characteristic distance for "change" in the exposure factor, and represents the distance over which the exposure associated with a given source will change by a factor of $1/e$. A large characteristic distance suggests that the exposure decreases slowly with distance resulting in a longer influence of each site on exposure and, potentially, influence of a greater number of sites on exposure at any given location. The distance decay function is illustrated in Fig. 1. Since the appropriate lag time between exposure to benzene and onset of lymphoma remains unknown [27,35], four separate exposure periods prior to and overlapping the case data time frame of 1999–2008 were also examined to identify the exposure period and scaling factor characteristics that best predicted NHL risk within our models: 1989–1998, 1989–1993, 1994–1998, 1994–2003.

3.2. Statistical models

Poisson regression is a commonly used approach for modeling the relationship between count data and a variable of interest, particularly for studies with smaller areal units and rare diseases [36]. Using Poisson regression, we model the number of NHL cases as a function of exposure level, while fitting an offset, the expected number of cases in each tract, determined by the age, sex, and race demographics of the population following our previously published methods [27]. In addition to accounting for age, race, and gender in our model, the expected count represents a measure of population in each census tract. Thus, even though census tracts are intended to be approximately the same population, we account for variability in tract population sizes. The quantity of interest is the standardized incidence ratio (SIR), the ratio between the observed number of cases in the tract based on GCCR data and the expected number of cases, which provides a measure of risk. Risk ratios were estimated by exponentiating the model parameter estimates. Additionally, we fit NHL risk as a function of exposure, while adjusting for percent of the population 25 years and older who are high school graduates and median income at the census tract level.

One drawback to a Poisson model is the restriction that the mean must equal the variance in the count distribution. An alternative approach for risk estimation allows for overdispersion (i.e., greater variability in the data than expected from a Poisson model) via negative binomial regression. Negative binomial models allow for maximum likelihood estimation of an additional shape parameter, which provides flexibility for estimating the variance as distinct from the mean. Because of the potential high variability in sample size and risk across census tracts, we constructed Poisson regression and negative binomial regression models and compared the goodness-of-fit for each approach. This analysis was also repeated for two NHL subtypes, diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL).

Table 1
Model goodness-of-fit statistics for prediction of non-Hodgkin lymphoma risk.

Model	Exposure period (Years)	Scaling factor (km)	Deviance/*df*	AIC
Negative binomial	1989–1993	4	1.16	7923.7
Negative binomial	1989–1993	8	1.16	7939.0
Negative binomial	1989–1993	16	1.16	7961.4
Negative binomial	1989–1993	24	1.17	7986.8
Negative binomial	1989–1998	4	1.17	7973.7
Negative binomial	1989–1998	8	1.17	7974.2
Negative binomial	1989–1998	16	1.17	8007.5
Negative binomial	1989–1998	24	1.18	8021.4
Negative binomial	1994–1998	4	1.17	7981.7
Negative binomial	1994–1998	8	1.17	7985.5
Negative binomial	1994–1998	16	1.17	8010.5
Negative binomial	1994–1998	24	1.16	7968.6
Negative binomial	1994–2003	4	1.17	7997.3
Negative binomial	1994–2003	8	1.17	7985.3
Negative binomial	1994–2003	16	1.18	7998.3
Negative binomial	1994–2003	24	1.18	8003.4
Poisson	1989–1993	4	1.60	8058.5
Poisson	1989–1993	8	1.61	8079.8
Poisson	1989–1993	16	1.63	8111.7
Poisson	1989–1993	24	1.65	8143.4
Poisson	1989–1998	4	1.64	8120.7
Poisson	1989–1998	8	1.63	8117.4
Poisson	1989–1998	16	1.67	8169.8
Poisson	1989–1998	24	1.68	8190.1
Poisson	1994–1998	4	1.64	8134.8
Poisson	1994–1998	8	1.65	8137.1
Poisson	1994–1998	16	1.67	8180.0
Poisson	1994–1998	24	1.64	8122.8
Poisson	1994–2003	4	1.66	8156.4
Poisson	1994–2003	8	1.65	8136.0
Poisson	1994–2003	16	1.65	8151.4
Poisson	1994–2003	24	1.66	8166.5

3.3. Goodness-of-fit criteria

Goodness-of-fit for both the Poisson and negative binomial models was assessed using the residual deviance divided by the degrees of freedom (df). AIC values were reported in order to compare the relative fits of the Poisson and negative binomial models at various scaling factors and exposure periods. Plots of observed vs. expected values provided additional assessment of model fit for each observation to check for high leverage and/or influential observations.

Risk ratios and confidence intervals (CI) were reported. Significance was assessed at the $p < 0.05$ level, and statistical analysis was performed using R 2.15.1 [37] (R Statistical Computing, Vienna, Austria). A Bonferroni approach was further explored to control for Type I error given the high number of hypothesis tests, with 128 total pairwise tests for both Poisson and negative binomial regression techniques combined. The R package glm.nb [38] was used for estimating negative binomial model parameters when the shape parameter was unknown. Census tract shapefiles were uploaded to R using the package maptools [39], and we observed and plotted the spatial distributions of benzene exposure levels along with the SIRs for NHL.

4. Results

11,323 NHL cases with available demographic information were geocoded across 1616 tracts in Georgia from 1999 to 2008, yielding an average of 7.0 NHL cases per tract (minimum: 0, 25th percentile: 3, median: 6, 75th percentile: 10, maximum: 47). Of the 22 benzene TRI release sites in Georgia from 1989 to 2003, 7 facilities reported benzene released from 1989 to 1993, 18 facilities reported benzene released from 1989 to 1998, 16 facilities reported benzene released from 1994 to 1998, and 19 facilities reported benzene released from 1994 to 2003. The average number of years that a facility reported benzene release from 1989 to 2003 was 6.2 years. Cumulative exposure levels were categorized into quintiles.

The map of observed SIRs for each Georgia census tract for NHL is shown in Fig. 2. Elevated risk was concentrated in the metro Atlanta area, defined as Fulton, DeKalb, Clayton, Cobb, and Gwinnett counties, as well as some rural census tracts, indicated by the darker shades. Previous analyses of NHL incidence based on these data using Moran's *I*, a measure of spatial autocorrelation, yielded evidence of significant global spatial correlation [27].

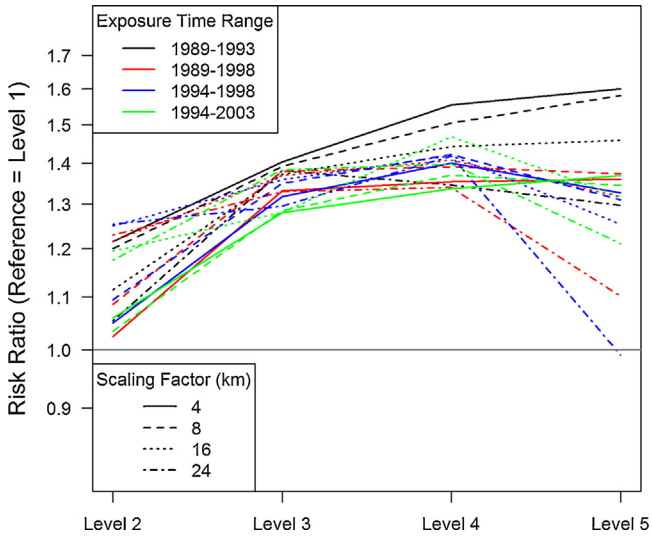

Fig. 3. Negative binomial model risk ratios, using Level 1 as the reference level. Lines are color-coded by exposure period: 1989–1993 (black), 1989–1998 (red), 1994–1998 (blue), 1994–2003 (green). Line type indicates scaling factor.

Table 2
Negative binomial regression statistics for the 1989–1993 and the 1989–1998 exposure periods.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1989–1993	4	Exposure: Level 2 vs. Level 1	[2.4e-12–3.5e-06] vs. [7.3e-33–2.4e-12]	1.22	(1.12, 1.32)	<0.0001
		Exposure: Level 3 vs. Level 1	[3.5e-06–1.1] vs. [7.3e-33–2.4e-12]	1.40	(1.30, 1.52)	<0.0001
		Exposure: Level 4 vs. Level 1	[1.1–160] vs. [7.3e-33–2.4e-12]	1.56	(1.44, 1.68)	<0.0001
		Exposure: Level 5 vs. Level 1	[160–360,000] vs. [7.3e-33–2.4e-12]	1.60	(1.48, 1.74)	<0.0001
1989–1993	8	Exposure: Level 2 vs. Level 1	[0.00015–0.15] vs. [6.9e-15–0.00015]	1.20	(1.11, 1.30)	<0.0001
		Exposure: Level 3 vs. Level 1	[0.15–78] vs. [6.9e-15–0.00015]	1.39	(1.29, 1.50)	<0.0001
		Exposure: Level 4 vs. Level 1	[78–710] vs. [6.9e-15–0.00015]	1.50	(1.39, 1.63)	<0.0001
		Exposure: Level 5 vs. Level 1	[710–570,000] vs. [6.9e-15–0.00015]	1.58	(1.46, 1.71)	<0.0001
1989–1993	16	Exposure: Level 2 vs. Level 1	[1.9–86] vs. [2.4e-05–1.9]	1.11	(1.03, 1.21)	0.0086
		Exposure: Level 3 vs. Level 1	[86–720] vs. [2.4e-05–1.9]	1.37	(1.27, 1.48)	<0.0001
		Exposure: Level 4 vs. Level 1	[720–2000] vs. [2.4e-05–1.9]	1.44	(1.34, 1.56)	<0.0001
		Exposure: Level 5 vs. Level 1	[2000–720,000] vs. [2.4e-05–1.9]	1.46	(1.35, 1.58)	<0.0001
1989–1993	24	Exposure: Level 2 vs. Level 1	[55–740] vs. [0.072–55]	1.05	(0.97, 1.14)	0.1997
		Exposure: Level 3 vs. Level 1	[740–2000] vs. [0.072–55]	1.38	(1.28, 1.49)	<0.0001
		Exposure: Level 4 vs. Level 1	[2000–3400] vs. [0.072–55]	1.35	(1.25, 1.45)	<0.0001
		Exposure: Level 5 vs. Level 1	[3400–770,000] vs. [0.072–55]	1.30	(1.20, 1.41)	<0.0001
1989–1998	4	Exposure: Level 2 vs. Level 1	[1.2e-06–0.0082] vs. [3.3e-17–1.2e-06]	1.02	(0.95, 1.11)	0.5591
		Exposure: Level 3 vs. Level 1	[0.0082–14] vs. [3.3e-17–1.2e-06]	1.33	(1.24, 1.43)	<0.0001
		Exposure: Level 4 vs. Level 1	[14–370] vs. [3.3e-17–1.2e-06]	1.35	(1.26, 1.46)	<0.0001
		Exposure: Level 5 vs. Level 1	[370–560,000] vs. [3.3e-17–1.2e-06]	1.36	(1.26, 1.47)	<0.0001
1989–1998	8	Exposure: Level 2 vs. Level 1	[0.056–7.6] vs. [8.7e-07–0.056]	1.09	(1.00, 1.17)	0.0408
		Exposure: Level 3 vs. Level 1	[7.6–280] vs. [8.7e-07–0.056]	1.38	(1.28, 1.48)	<0.0001
		Exposure: Level 4 vs. Level 1	[280–1400] vs. [8.7e-07–0.056]	1.39	(1.29, 1.50)	<0.0001
		Exposure: Level 5 vs. Level 1	[1400–890,000] vs. [8.7e-07–0.056]	1.37	(1.27, 1.49)	<0.0001
1989–1998	16	Exposure: Level 2 vs. Level 1	[39–350] vs. [0.12–39]	1.22	(1.13, 1.31)	<0.0001
		Exposure: Level 3 vs. Level 1	[350–1700] vs. [0.12–39]	1.37	(1.27, 1.48)	<0.0001
		Exposure: Level 4 vs. Level 1	[1700–3500] vs. [0.12–39]	1.41	(1.30, 1.52)	<0.0001
		Exposure: Level 5 vs. Level 1	[3500–1,100,000] vs. [0.12–39]	1.32	(1.22, 1.43)	<0.0001
1989–1998	24	Exposure: Level 2 vs. Level 1	[530–2300] vs. [4.2–530]	1.23	(1.14, 1.33)	<0.0001
		Exposure: Level 3 vs. Level 1	[2300–3900] vs. [4.2–530]	1.33	(1.23, 1.43)	<0.0001
		Exposure: Level 4 vs. Level 1	[3900–5600] vs. [4.2–530]	1.34	(1.24, 1.45)	<0.0001
		Exposure: Level 5 vs. Level 1	[5600–1,200,000] vs. [4.2–530]	1.10	(1.02, 1.19)	0.0211

A total of 32 models were run, with both Poisson and negative binomial parameterizations used for 4 different spatial scaling factors and 4 different time frames for exposure accumulation. Model fit criteria are displayed in Table 1. Based on the deviance/d*f* criteria for goodness-of-fit, the Poisson models demonstrated poor fit. The negative binomial models demonstrated much better fit with all deviance/d*f* values near 1. Consequently, the following results are drawn from the negative binomial models.

Risk ratios for all negative binomial models are displayed in Fig. 3, using the lowest exposure level as the reference. In nearly every model, lymphoma risk increased as the level of exposure increased, particularly for the three highest levels of exposure in comparison to the lowest level. Sixty-three of the 64 risk ratios (from 16 total models) comparing an upper level of exposure to the lowest level were above the null value of 1. Fifty-eight of the 64 95% CIs did not contain the null value of 1, and of the remaining 6, 5 resulted from a comparison of Level 2 (the 2nd lowest level) vs. Level 1 (Tables 2 and 3). Using a conservative Bonferroni approach for allocating Type I error, 54 of the 64 hypothesis tests yielded *p*-values below 0.05/128 (threshold = 0.00039).

The exposure period of 1989–1993 with the smallest scaling factor of 4 km yielded the lowest AIC value for NHL. Moreover, a scaling factor of 4 km provided the lowest AIC value for the 1989–1998 exposure period, while scaling factors of 24 km and 8 km produced the lowest AIC values for the 1994–1998 and 1994–2003 exposure periods, respectively. For the scaling factors of 4 km, 8 km, and 16 km, the exposure period of 1989–1993 produced the lowest AIC values, while for the scaling factor of 24 km, the 1994–1998 exposure period produced the lowest AIC value. Thus, the exposure periods that included the oldest exposure data were better fit with smaller scaling factors, yielding a stronger local effect. The exposure periods including the more recent exposure data were better fit with larger scaling factors, yielding a weaker local effect and a stronger regional effect.

The map of exposure levels for the best fitting model in relation to location of benzene release sites in that time frame is displayed in Fig. 4. The highest exposure level census tracts were observed in the center of metro Atlanta as well as in close proximity to the Augusta, GA, benzene release site. Risk ratios, 95% CIs, and *p*-values for this model are reported in Table 2. In this model, a Level 4 tract had a 56% higher risk of NHL than a Level 1 tract; likewise, risk was 60% higher in a Level 5 tract vs. a Level 1 tract. In the best fitting model, when adjusting for SES census tract level variables such as percent of the population 25 years and older who are high school graduates and median income, the results remain consistent where higher exposure levels have higher risk of NHL compared to the lowest levels (Table 4). For the 4 km scaling factor and exposure time frame of 1989–1993, the cut points separating exposure levels 1–5 were: 2.4×10^{-12} kg, 3.5×10^{-6} kg, 1.1 kg, and 160 kg, with a minimum exposure of 7.3×10^{-33} kg and a maximum of 360,000 kg. Results for the subtype analysis are reported in Appendix.

Table 3
Negative binomial regression statistics for the 1994–2003 and the 1994–1998 exposure periods.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1994–1998	4	Exposure: Level 2 vs. Level 1	[8.2e-07–0.0038] vs. [3.3e-17–8.2e-07]	1.05	(0.97, 1.13)	0.2378
		Exposure: Level 3 vs. Level 1	[0.0038–6.4] vs. [3.3e-17–8.2e-07]	1.32	(1.23, 1.42)	<0.0001
		Exposure: Level 4 vs. Level 1	[6.4–70] vs. [3.3e-17–8.2e-07]	1.40	(1.30, 1.51)	<0.0001
		Exposure: Level 5 vs. Level 1	[70–200,000] vs. [3.3e-17–8.2e-07]	1.33	(1.23, 1.43)	<0.0001
1994–1998	8	Exposure: Level 2 vs. Level 1	[0.037–3.8] vs. [8.7e-07–0.037]	1.09	(1.01, 1.18)	0.0243
		Exposure: Level 3 vs. Level 1	[3.8–98] vs. [8.7e-07–0.037]	1.35	(1.26, 1.45)	<0.0001
		Exposure: Level 4 vs. Level 1	[98–360] vs. [8.7e-07–0.037]	1.42	(1.32, 1.53)	<0.0001
		Exposure: Level 5 vs. Level 1	[360–320,000] vs. [8.7e-07–0.037]	1.31	(1.21, 1.41)	<0.0001
1994–1998	16	Exposure: Level 2 vs. Level 1	[23–160] vs. [0.12–23]	1.25	(1.16, 1.35)	<0.0001
		Exposure: Level 3 vs. Level 1	[160–630] vs. [0.12–23]	1.36	(1.26, 1.46)	<0.0001
		Exposure: Level 4 vs. Level 1	[630–1100] vs. [0.12–23]	1.42	(1.31, 1.53)	<0.0001
		Exposure: Level 5 vs. Level 1	[1100–410,000] vs. [0.12–23]	1.25	(1.16, 1.36)	<0.0001
1994–1998	24	Exposure: Level 2 vs. Level 1	[290–1100] vs. [2.2–290]	1.25	(1.17, 1.35)	<0.0001
		Exposure: Level 3 vs. Level 1	[1100–1600] vs. [2.2–290]	1.30	(1.20, 1.39)	<0.0001
		Exposure: Level 4 vs. Level 1	[1600–1900] vs. [2.2–290]	1.40	(1.32, 1.53)	<0.0001
		Exposure: Level 5 vs. Level 1	[1900–440,000] vs. [2.2–290]	0.99	(0.92, 1.07)	0.8065
1994–2003	4	Exposure: Level 2 vs. Level 1	[9.8e-06–0.018] vs. [2e-13–9.8e-06]	1.06	(0.98, 1.15)	0.1532
		Exposure: Level 3 vs. Level 1	[0.018–19] vs. [2e-13–9.8e-06]	1.28	(1.19, 1.38)	<0.0001
		Exposure: Level 4 vs. Level 1	[19–250] vs. [2e-13–9.8e-06]	1.34	(1.24, 1.44)	<0.0001
		Exposure: Level 5 vs. Level 1	[250–230,000] vs. [2e-13–9.8e-06]	1.37	(1.27, 1.48)	<0.0001
1994–2003	8	Exposure: Level 2 vs. Level 1	[0.17–11] vs. [2.9e-05–0.17]	1.03	(0.96, 1.12)	0.3982
		Exposure: Level 3 vs. Level 1	[11–270] vs. [2.9e-05–0.17]	1.29	(1.19, 1.38)	<0.0001
		Exposure: Level 4 vs. Level 1	[270–1300] vs. [2.9e-05–0.17]	1.37	(1.27, 1.48)	<0.0001
		Exposure: Level 5 vs. Level 1	[1300–370,000] vs. [2.9e-05–0.17]	1.35	(1.25, 1.45)	<0.0001
1994–2003	16	Exposure: Level 2 vs. Level 1	[56–370] vs. [0.17–56]	1.19	(1.11, 1.29)	<0.0001
		Exposure: Level 3 vs. Level 1	[370–1600] vs. [0.17–56]	1.28	(1.19, 1.38)	<0.0001
		Exposure: Level 4 vs. Level 1	[1600–3900] vs. [0.17–56]	1.47	(1.36, 1.58)	<0.0001
		Exposure: Level 5 vs. Level 1	[3900–460,000] vs. [0.17–56]	1.32	(1.22, 1.42)	<0.0001
1994–2003	24	Exposure: Level 2 vs. Level 1	[570–1800] vs. [3.9–570]	1.18	(1.09, 1.27)	<0.0001
		Exposure: Level 3 vs. Level 1	[1800–4100] vs. [3.9–570]	1.38	(1.29, 1.49)	<0.0001
		Exposure: Level 4 vs. Level 1	[4100–6200] vs. [3.9–570]	1.40	(1.30, 1.51)	<0.0001
		Exposure: Level 5 vs. Level 1	[6200–500,000] vs. [3.9–570]	1.21	(1.12, 1.31)	<0.0001

5. Discussion

This study extends our existing understanding of the relationship between proximity to benzene release sites and NHL risk by utilizing TRI data to weight our measure of exposure by both proximity to release sites and amount of release. In addition, we utilized a method for estimating the level of benzene exposure in a given geographic space and fit statistical models to examine associations between exposure periods, lag times, and lymphoma incident cases. For all models, we evaluated the goodness-of-fit and assessed the optimal scaling factor parameter to measure exposure in addition to the effects of time lag variations on goodness-of-fit.

Across our models, we consistently found that census tracts that were considered higher-exposure zones exhibited higher risk of NHL than lower-exposure zones, and that a statistically significant effect was noted even at very low exposure levels—far below occupational exposure levels. This consistency across conditions suggests that the effect of passive benzene exposure on lymphoma risk was independent of both time lag and scaling factor. Although the results are similar to our previous findings [27], the added magnitude component strengthens the argument that passive exposure is associated with NHL risk, not only as a result of distance from release sites, but also as a function of amount of benzene released from these facilities over time.

Other studies have demonstrated that occupational exposure to benzene even in low doses increases cancer risk [2,3]. In addition,

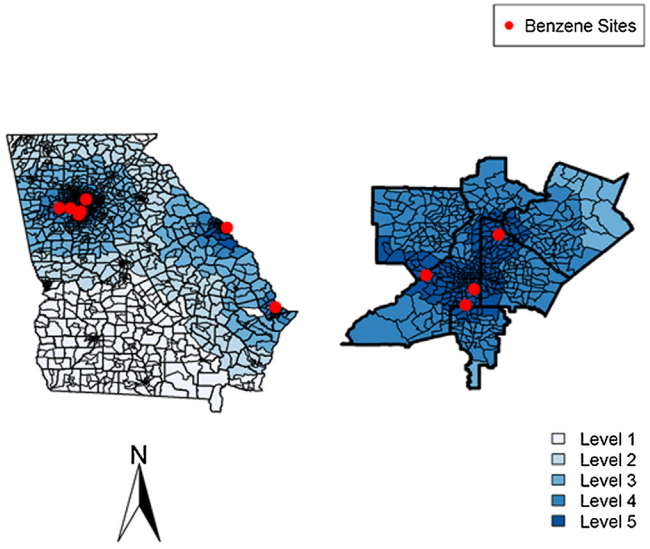

Fig. 4. Locations of benzene release sites in Georgia and exposure levels from 1989 to 1993, with a scaling factor of 4 km. This exposure period and scaling factor produced the lowest AIC value. Cumulative exposure was categorized into 5 levels based on quintiles. Census tracts are color-coded according to exposure level, where the darkest blue represents the highest level of exposure.

Table 4
Negative binomial regression statistics for the 1989–1993 exposure period and 4-km scaling factor adjusting for percent of the population 25 years and older who are high school graduates and median income.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1989–1993	4	Exposure: Level 2 vs. Level 1	[8.2e-07–0.0038] vs. [3.3e-17–8.2e-07]	1.18	(1.04, 1.33)	0.0124
		Exposure: Level 3 vs. Level 1	[0.0038–6.4] vs. [3.3e-17–8.2e-07]	1.30	(1.15, 1.47)	<0.0001
		Exposure: Level 4 vs. Level 1	[6.4–70] vs. [3.3e-17–8.2e-07]	1.27	(1.12, 1.45)	0.0004
		Exposure: Level 5 vs. Level 1	[70–200,000] vs. [3.3e-17–8.2e-07]	1.50	(1.32, 1.71)	<0.0001

our group performed the first population-based analysis to demonstrate that passive exposure through residential proximity to EPA-identified benzene release sites is associated with increased risk of NHL [27]. In this current study, we used amount of release and coordinate data extracted from the EPA's TRI to expand upon this finding and utilized a method for quantifying the level of passive benzene exposure based on proximity to each site and amount of release. We further tested this method using varying spatial scaling factors and exposure periods to assess for different assumed time lags between exposure and disease. It is not known how quickly or broadly benzene dissipates in the atmosphere from these release sites, nor is it known exactly how quickly lymphoma develops in the setting of long-term passive exposure to benzene. As a result, using several scaling factors and lag times enabled us to assess how consistent our results were under reasonable, varying conditions.

Limitations of this study include the use of retrospective data to assess the impact of benzene exposure on cancer risk. Despite the use of varying lag times for exposure and onset of cases, our results indicate only an association between exposure and disease, rather than a causative effect. In addition, data were aggregated at the census tract level in order to capture the spatial association between proximity to release sites and cancer risk; as a result, we should be cautious in extrapolating these findings to the individual level, since environmental exposure does not necessarily equate to individual exposure. Furthermore, by aggregating to the census tract level, we encountered census tracts of various geographic sizes, which could lead to worse exposure classification in larger sized tracts with smaller populations when utilizing smaller scaling factors. Additionally, other sources of benzene such as occupational and traffic sources were not characterized in this aggregated analysis; however, this type of analysis was exploratory and provides hypothesis-generating results for subsequent studies involving patient-level data.

We also recognize that the exposure cut points for the best fitting model were very small (<1 kg for the 20th and 40th percentile), and it could be argued that lower exposure levels for smaller scaling factors should be grouped. However, for the purposes of consistency across scaling factors and exposure periods, we kept the number of quantiles at 5 for all models. It also should be noted that the risk ratio between Levels 1 and 2 exposure levels was not statistically significant for several scaling factors and exposure periods, which further supports a strategy of grouping exposure levels with small differences in cumulative exposure.

Future analyses could include more advanced statistical models for isolating the effect of benzene exposure on cancer risk, such as spatial Bayesian hierarchical models with conditionally autoregressive random effects [40–42]. These types of models attempt to classify observed correlated data, such as spatial data, and control for other potential unmeasured confounding risk factors. Such models also provide more precise incidence estimates for small areas with lower observed case numbers, which often yield inflated rate and ratio estimates if left unadjusted.

Despite the use of several spatial scaling factors in our models as a measure of model sensitivity, we could not pinpoint the precise dispersion of benzene from these sites over time, and we found that only 9.5% of census tracts remained in the same exposure level across all 4 scaling factors and 4 exposure periods. Thus, future work could include identifying more precisely the benzene dispersion/scaling parameter from a release site. In order to isolate the true exposure effect on disease risk in an aggregated spatial analysis and to reduce the spatial uncertainty from varying scaling factors, it is necessary to estimate the rate at which benzene diffuses into the atmosphere surrounding these sites. It is likely that the diffusion would be influenced by meteorological data, such as wind speeds and direction. Collecting observational benzene exposure data around release sites at varying distances over time in order to estimate the benzene dispersion parameter may further reduce spatial uncertainty in estimations of exposure using this approach. Subsequent investigation of the effects of passive benzene exposure on cancer risk should also involve individual patient-level data in the form of a case-control study or other longitudinal observational study where benzene levels can be monitored at the individual exposure level. We are engaged in ongoing studies measuring BTEX releases at various distances around TRI sites and assessing patient-level risk factors for individuals with NHL in Georgia. Upcoming goals of our research in this area aim to improve our understanding of the interactions between long-term passive exposure to VOCs and clinical, lifestyle, and genetic factors that contribute to cancer risk for individuals and populations. Our data provide a foundation for this approach.

Conflicts of interest

None.

Authorship contribution

JS wrote the manuscript and performed the data analysis. CB collected data, and reviewed and edited the manuscript. KW provided the cancer data. JK, AB, PR, and LW reviewed and edited the manuscript. CF developed the project, advised and mentored JS, and reviewed and edited the manuscript.

Acknowledgements

Research reported in this publication was supported in part by Dr. Switchenko's Pam Rollins Research Award, a philanthropic award provided by the Winship Cancer Institute of Emory University, by Dr. Flowers' National Cancer InstituteR21CA158686, and by the NIH/NCI under award number P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix.

See Table A1a, Table A1b, Table A2a, Table A2b, Table A3a, Table A3b.

Table A1a
Model goodness-of-fit statistics for prediction of DLBCL risk.

Model	Exposure period (Years)	Scaling factor (km)	Deviance/df	AIC
Negative binomial	1989–1993	4	1.13	5656.1
Negative binomial	1989–1993	8	1.14	5657.9
Negative binomial	1989–1993	16	1.14	5674.8
Negative binomial	1989–1993	24	1.14	5683.0
Negative binomial	1989–1998	4	1.14	5674.2
Negative binomial	1989–1998	8	1.14	5671.6
Negative binomial	1989–1998	16	1.14	5687.1
Negative binomial	1989–1998	24	1.14	5700.2
Negative binomial	1994–1998	4	1.14	5683.5
Negative binomial	1994–1998	8	1.14	5683.3
Negative binomial	1994–1998	16	1.14	5683.6
Negative binomial	1994–1998	24	1.13	5666.1
Negative binomial	1994–2003	4	1.14	5682.6
Negative binomial	1994–2003	8	1.14	5678.1
Negative binomial	1994–2003	16	1.15	5681.8
Negative binomial	1994–2003	24	1.14	5681.4
Poisson	1989–1993	4	1.31	5680.7
Poisson	1989–1993	8	1.31	5682.3
Poisson	1989–1993	16	1.32	5701.9
Poisson	1989–1993	24	1.33	5711.1
Poisson	1989–1998	4	1.32	5700.0
Poisson	1989–1998	8	1.32	5696.5
Poisson	1989–1998	16	1.33	5714.8
Poisson	1989–1998	24	1.34	5730.4
Poisson	1994–1998	4	1.33	5710.9
Poisson	1994–1998	8	1.33	5710.0
Poisson	1994–1998	16	1.33	5711.4
Poisson	1994–1998	24	1.32	5693.2
Poisson	1994–2003	4	1.33	5709.7
Poisson	1994–2003	8	1.33	5703.3
Poisson	1994–2003	16	1.33	5706.0
Poisson	1994–2003	24	1.33	5707.3

Table A1b
Model goodness-of-fit statistics for prediction of FL risk.

Model	Exposure period (Years)	Scaling factor (km)	Deviance/df	AIC
Negative binomial	1989–1993	4	1.05	4359.3
Negative binomial	1989–1993	8	1.05	4364.1
Negative binomial	1989–1993	16	1.05	4358.4
Negative binomial	1989–1993	24	1.05	4364.0
Negative binomial	1989–1998	4	1.06	4369.7
Negative binomial	1989–1998	8	1.06	4357.1
Negative binomial	1989–1998	16	1.05	4362.2
Negative binomial	1989–1998	24	1.05	4363.9
Negative binomial	1994–1998	4	1.06	4376.5
Negative binomial	1994–1998	8	1.06	4374.7
Negative binomial	1994–1998	16	1.05	4373.6
Negative binomial	1994–1998	24	1.05	4347.9
Negative binomial	1994–2003	4	1.06	4382.6
Negative binomial	1994–2003	8	1.06	4381.3
Negative binomial	1994–2003	16	1.06	4373.7
Negative binomial	1994–2003	24	1.05	4377.2
Poisson	1989–1993	4	1.12	4362.8
Poisson	1989–1993	8	1.12	4367.9
Poisson	1989–1993	16	1.12	4361.5
Poisson	1989–1993	24	1.12	4367.5
Poisson	1989–1998	4	1.13	4373.7
Poisson	1989–1998	8	1.12	4359.7
Poisson	1989–1998	16	1.12	4365.8
Poisson	1989–1998	24	1.12	4367.5
Poisson	1994–1998	4	1.13	4380.9
Poisson	1994–1998	8	1.13	4379.1
Poisson	1994–1998	16	1.13	4378.5
Poisson	1994–1998	24	1.11	4350.1
Poisson	1994–2003	4	1.14	4387.7
Poisson	1994–2003	8	1.13	4385.6
Poisson	1994–2003	16	1.13	4377.8
Poisson	1994–2003	24	1.13	4382.2

Table A2a
Negative binomial regression statistics for the 1989–1993 and the 1989–1998 exposure periods – DLBCL.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1989–1993	4	Exposure: Level 2 vs. Level 1	[2.4e-12–3.5e-06] vs. [7.3e-33–2.4e-12]	1.18	(1.04, 1.34)	0.0099
		Exposure: Level 3 vs. Level 1	[3.5e-06–1.1] vs. [7.3e-33–2.4e-12]	1.36	(1.20, 1.53)	<0.0001
		Exposure: Level 4 vs. Level 1	[1.1–160] vs. [7.3e-33–2.4e-12]	1.49	(1.32, 1.67)	<0.0001
		Exposure: Level 5 vs. Level 1	[160–360,000] vs. [7.3e-33–2.4e-12]	1.66	(1.46, 1.88)	<0.0001
1989–1993	8	Exposure: Level 2 vs. Level 1	[0.00015–0.15] vs. [6.9e-15–0.00015]	1.21	(1.06, 1.37)	0.0035
		Exposure: Level 3 vs. Level 1	[0.15–78] vs. [6.9e-15–0.00015]	1.37	(1.21, 1.54)	<0.0001
		Exposure: Level 4 vs. Level 1	[78–710] vs. [6.9e-15–0.00015]	1.51	(1.34, 1.70)	<0.0001
		Exposure: Level 5 vs. Level 1	[710–570,000] vs. [6.9e-15–0.00015]	1.65	(1.46, 1.87)	<0.0001
1989–1993	16	Exposure: Level 2 vs. Level 1	[1.9–86] vs. [2.4e-05–1.9]	1.09	(0.96, 1.23)	0.1892
		Exposure: Level 3 vs. Level 1	[86–720] vs. [2.4e-05–1.9]	1.33	(1.18, 1.49)	<0.0001
		Exposure: Level 4 vs. Level 1	[720–2000] vs. [2.4e-05–1.9]	1.38	(1.23, 1.56)	<0.0001
		Exposure: Level 5 vs. Level 1	[2000–720,000] vs. [2.4e-05–1.9]	1.49	(1.32, 1.69)	<0.0001
1989–1993	24	Exposure: Level 2 vs. Level 1	[55–740] vs. [0.072–55]	1.01	(0.89, 1.14)	0.9361
		Exposure: Level 3 vs. Level 1	[740–2000] vs. [0.072–55]	1.34	(1.19, 1.50)	<0.0001
		Exposure: Level 4 vs. Level 1	[2000–3400] vs. [0.072–55]	1.31	(1.16, 1.47)	<0.0001
		Exposure: Level 5 vs. Level 1	[3400–770,000] vs. [0.072–55]	1.32	(1.17, 1.49)	<0.0001
1989–1998	4	Exposure: Level 2 vs. Level 1	[1.2e-06–0.0082] vs. [3.3e-17–1.2e-06]	1.00	(0.89, 1.13)	0.9589
		Exposure: Level 3 vs. Level 1	[0.0082–14] vs. [3.3e-17–1.2e-06]	1.28	(1.14, 1.43)	<0.0001
		Exposure: Level 4 vs. Level 1	[14–370] vs. [3.3e-17–1.2e-06]	1.32	(1.18, 1.48)	<0.0001
		Exposure: Level 5 vs. Level 1	[370–560,000] vs. [3.3e-17–1.2e-06]	1.44	(1.28, 1.62)	<0.0001
1989–1998	8	Exposure: Level 2 vs. Level 1	[0.056–7.6] vs. [8.7e-07–0.056]	1.13	(1.00, 1.27)	0.0483
		Exposure: Level 3 vs. Level 1	[7.6–280] vs. [8.7e-07–0.056]	1.36	(1.21, 1.52)	<0.0001
		Exposure: Level 4 vs. Level 1	[280–1400] vs. [8.7e-07–0.056]	1.41	(1.26, 1.58)	<0.0001
		Exposure: Level 5 vs. Level 1	[1400–890,000] vs. [8.7e-07–0.056]	1.52	(1.35, 1.71)	<0.0001
1989–1998	16	Exposure: Level 2 vs. Level 1	[39–350] vs. [0.12–39]	1.26	(1.12, 1.42)	0.0002
		Exposure: Level 3 vs. Level 1	[350–1700] vs. [0.12–39]	1.35	(1.21, 1.52)	<0.0001
		Exposure: Level 4 vs. Level 1	[1700–3500] vs. [0.12–39]	1.43	(1.28, 1.61)	<0.0001
		Exposure: Level 5 vs. Level 1	[3500–1,100,000] vs. [0.12–39]	1.45	(1.28, 1.63)	<0.0001
1989–1998	24	Exposure: Level 2 vs. Level 1	[530–2300] vs. [4.2–530]	1.26	(1.13, 1.41)	0.0001
		Exposure: Level 3 vs. Level 1	[2300–3900] vs. [4.2–530]	1.32	(1.18, 1.48)	<0.0001
		Exposure: Level 4 vs. Level 1	[3900–5600] vs. [4.2–530]	1.37	(1.23, 1.54)	<0.0001
		Exposure: Level 5 vs. Level 1	[5600–1,200,000] vs. [4.2–530]	1.15	(1.02, 1.31)	0.0239

Table A2b
Negative binomial regression statistics for the 1989–1993 and the 1989–1998 exposure periods – FL.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1989–1993	4	Exposure: Level 2 vs. Level 1	[2.4e-12–3.5e-06] vs. [7.3e-33–2.4e-12]	1.29	(1.10, 1.51)	0.0016
		Exposure: Level 3 vs. Level 1	[3.5e-06–1.1] vs. [7.3e-33–2.4e-12]	1.41	(1.21, 1.65)	<0.0001
		Exposure: Level 4 vs. Level 1	[1.1–160] vs. [7.3e-33–2.4e-12]	1.57	(1.35, 1.83)	<0.0001
		Exposure: Level 5 vs. Level 1	[160–360,000] vs. [7.3e-33–2.4e-12]	1.22	(1.03, 1.45)	0.022
1989–1993	8	Exposure: Level 2 vs. Level 1	[0.00015–0.15] vs. [6.9e-15–0.00015]	1.25	(1.07, 1.47)	0.0048
		Exposure: Level 3 vs. Level 1	[0.15–78] vs. [6.9e-15–0.00015]	1.38	(1.19, 1.61)	<0.0001
		Exposure: Level 4 vs. Level 1	[78–710] vs. [6.9e-15–0.00015]	1.51	(1.30, 1.76)	<0.0001
		Exposure: Level 5 vs. Level 1	[710–570,000] vs. [6.9e-15–0.00015]	1.17	(0.99, 1.39)	0.0611
1989–1993	16	Exposure: Level 2 vs. Level 1	[1.9–86] vs. [2.4e-05–1.9]	1.22	(1.05, 1.43)	0.0118
		Exposure: Level 3 vs. Level 1	[86–720] vs. [2.4e-05–1.9]	1.45	(1.25, 1.68)	<0.0001
		Exposure: Level 4 vs. Level 1	[720–2000] vs. [2.4e-05–1.9]	1.50	(1.29, 1.75)	<0.0001
		Exposure: Level 5 vs. Level 1	[2000–720,000] vs. [2.4e-05–1.9]	1.14	(0.97, 1.35)	0.1194
1989–1993	24	Exposure: Level 2 vs. Level 1	[55–740] vs. [0.072–55]	1.17	(1.01, 1.36)	0.0378
		Exposure: Level 3 vs. Level 1	[740–2000] vs. [0.072–55]	1.39	(1.21, 1.61)	<0.0001
		Exposure: Level 4 vs. Level 1	[2000–3400] vs. [0.072–55]	1.38	(1.19, 1.60)	<0.0001
		Exposure: Level 5 vs. Level 1	[3400–770,000] vs. [0.072–55]	1.03	(0.88, 1.22)	0.6924
1989–1998	4	Exposure: Level 2 vs. Level 1	[1.2e-06–0.0082] vs. [3.3e-17–1.2e-06]	0.95	(0.82, 1.10)	0.4982
		Exposure: Level 3 vs. Level 1	[0.0082–14] vs. [3.3e-17–1.2e-06]	1.25	(1.10, 1.43)	0.0009
		Exposure: Level 4 vs. Level 1	[14–370] vs. [3.3e-17–1.2e-06]	1.23	(1.07, 1.41)	0.0029
		Exposure: Level 5 vs. Level 1	[370–560,000] vs. [3.3e-17–1.2e-06]	0.96	(0.82, 1.13)	0.6264
1989–1998	8	Exposure: Level 2 vs. Level 1	[0.056–7.6] vs. [8.7e-07–0.056]	1.07	(0.93, 1.24)	0.3576
		Exposure: Level 3 vs. Level 1	[7.6–280] vs. [8.7e-07–0.056]	1.33	(1.16, 1.51)	<0.0001
		Exposure: Level 4 vs. Level 1	[280–1400] vs. [8.7e-07–0.056]	1.32	(1.15, 1.52)	0.0001
		Exposure: Level 5 vs. Level 1	[1400–890,000] vs. [8.7e-07–0.056]	0.91	(0.77, 1.07)	0.2623
1989–1998	16	Exposure: Level 2 vs. Level 1	[39–350] vs. [0.12–39]	1.22	(1.06, 1.40)	0.0067
		Exposure: Level 3 vs. Level 1	[350–1700] vs. [0.12–39]	1.29	(1.12, 1.47)	0.0003
		Exposure: Level 4 vs. Level 1	[1700–3500] vs. [0.12–39]	1.33	(1.15, 1.53)	0.0001
		Exposure: Level 5 vs. Level 1	[3500–1,100,000] vs. [0.12–39]	0.91	(0.77, 1.07)	0.2506
1989–1998	24	Exposure: Level 2 vs. Level 1	[530–2300] vs. [4.2–530]	1.24	(1.08, 1.42)	0.0018
		Exposure: Level 3 vs. Level 1	[2300–3900] vs. [4.2–530]	1.24	(1.08, 1.42)	0.0023
		Exposure: Level 4 vs. Level 1	[3900–5600] vs. [4.2–530]	1.17	(1.02, 1.35)	0.0283
		Exposure: Level 5 vs. Level 1	[5600–1,200,000] vs. [4.2–530]	0.84	(0.71, 0.99)	0.039

Table A3a
Negative binomial regression statistics for the 1994–2003 and the 1994–1998 exposure periods - DLBCL.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1994–1998	4	Exposure: Level 2 vs. Level 1	[8.2e-07–0.0038] vs. [3.3e-17–8.2e-07]	1.03	(0.91, 1.16)	0.6273
		Exposure: Level 3 vs. Level 1	[0.0038–6.4] vs. [3.3e-17–8.2e-07]	1.28	(1.14, 1.43)	<0.0001
		Exposure: Level 4 vs. Level 1	[6.4–70] vs. [3.3e-17–8.2e-07]	1.38	(1.23, 1.55)	<0.0001
		Exposure: Level 5 vs. Level 1	[70–200,000] vs. [3.3e-17–8.2e-07]	1.34	(1.20, 1.51)	<0.0001
1994–1998	8	Exposure: Level 2 vs. Level 1	[0.037–3.8] vs. [8.7e-07–0.037]	1.12	(1.00, 1.27)	0.0546
		Exposure: Level 3 vs. Level 1	[3.8–98] vs. [8.7e-07–0.037]	1.33	(1.19, 1.49)	<0.0001
		Exposure: Level 4 vs. Level 1	[98–360] vs. [8.7e-07–0.037]	1.44	(1.28, 1.62)	<0.0001
		Exposure: Level 5 vs. Level 1	[360–320,000] vs. [8.7e-07–0.037]	1.37	(1.22, 1.54)	<0.0001
1994–1998	16	Exposure: Level 2 vs. Level 1	[23–160] vs. [0.12–23]	1.29	(1.15, 1.45)	<0.0001
		Exposure: Level 3 vs. Level 1	[160–630] vs. [0.12–23]	1.35	(1.20, 1.51)	<0.0001
		Exposure: Level 4 vs. Level 1	[630–1100] vs. [0.12–23]	1.53	(1.36, 1.71)	<0.0001
		Exposure: Level 5 vs. Level 1	[1100–410,000] vs. [0.12–23]	1.29	(1.15, 1.45)	<0.0001
1994–1998	24	Exposure: Level 2 vs. Level 1	[290–1100] vs. [2.2–290]	1.28	(1.15, 1.43)	<0.0001
		Exposure: Level 3 vs. Level 1	[1100–1600] vs. [2.2–290]	1.32	(1.18, 1.48)	<0.0001
		Exposure: Level 4 vs. Level 1	[1600–1900] vs. [2.2–290]	1.49	(1.33, 1.67)	<0.0001
		Exposure: Level 5 vs. Level 1	[1900–440,000] vs. [2.2–290]	1.01	(0.90, 1.14)	0.8207
1994–2003	4	Exposure: Level 2 vs. Level 1	[9.8e-06–0.018] vs. [2e-13–9.8e-06]	1.02	(0.90, 1.15)	0.7864
		Exposure: Level 3 vs. Level 1	[0.018–19] vs. [2e-13–9.8e-06]	1.24	(1.11, 1.39)	0.0001
		Exposure: Level 4 vs. Level 1	[19–250] vs. [2e-13–9.8e-06]	1.31	(1.17, 1.47)	<0.0001
		Exposure: Level 5 vs. Level 1	[250–230,000] vs. [2e-13–9.8e-06]	1.41	(1.25, 1.58)	<0.0001
1994–2003	8	Exposure: Level 2 vs. Level 1	[0.17–11] vs. [2.9e-05–0.17]	1.07	(0.95, 1.21)	0.2475
		Exposure: Level 3 vs. Level 1	[11–270] vs. [2.9e-05–0.17]	1.26	(1.12, 1.40)	0.0001
		Exposure: Level 4 vs. Level 1	[270–1300] vs. [2.9e-05–0.17]	1.42	(1.27, 1.60)	<0.0001
		Exposure: Level 5 vs. Level 1	[1300–370,000] vs. [2.9e-05–0.17]	1.41	(1.26, 1.58)	<0.0001
1994–2003	16	Exposure: Level 2 vs. Level 1	[56–370] vs. [0.17–56]	1.22	(1.09, 1.38)	0.0009
		Exposure: Level 3 vs. Level 1	[370–1600] vs. [0.17–56]	1.28	(1.14, 1.44)	<0.0001
		Exposure: Level 4 vs. Level 1	[1600–3900] vs. [0.17–56]	1.51	(1.35, 1.70)	<0.0001
		Exposure: Level 5 vs. Level 1	[3900–460,000] vs. [0.17–56]	1.37	(1.22, 1.54)	<0.0001
1994–2003	24	Exposure: Level 2 vs. Level 1	[570–1800] vs. [3.9–570]	1.21	(1.08, 1.36)	0.0015
		Exposure: Level 3 vs. Level 1	[1800–4100] vs. [3.9–570]	1.40	(1.26, 1.57)	<0.0001
		Exposure: Level 4 vs. Level 1	[4100–6200] vs. [3.9–570]	1.49	(1.33, 1.67)	<0.0001
		Exposure: Level 5 vs. Level 1	[6200–500,000] vs. [3.9–570]	1.25	(1.11, 1.41)	0.0002

Table A3b
Negative binomial regression statistics for the 1994–2003 and the 1994–1998 exposure periods - FL.

Exposure period (Years)	Scaling factor (km)	Effect	Level (in kg)	Risk ratio	95% CI	*p*-value
1994–1998	4	Exposure: Level 2 vs. Level 1	[8.2e-07–0.0038] vs. [3.3e-17–8.2e-07]	0.97	(0.83, 1.12)	0.6477
		Exposure: Level 3 vs. Level 1	[0.0038–6.4] vs. [3.3e-17–8.2e-07]	1.22	(1.07, 1.40)	0.0028
		Exposure: Level 4 vs. Level 1	[6.4–70] vs. [3.3e-17–8.2e-07]	1.26	(1.09, 1.45)	0.0018
		Exposure: Level 5 vs. Level 1	[70–200,000] vs. [3.3e-17–8.2e-07]	1.03	(0.89, 1.20)	0.6756
1994–1998	8	Exposure: Level 2 vs. Level 1	[0.037–3.8] vs. [8.7e-07–0.037]	1.08	(0.94, 1.25)	0.2844
		Exposure: Level 3 vs. Level 1	[3.8–98] vs. [8.7e-07–0.037]	1.26	(1.10, 1.44)	0.0009
		Exposure: Level 4 vs. Level 1	[98–360] vs. [8.7e-07–0.037]	1.34	(1.17, 1.55)	<0.0001
		Exposure: Level 5 vs. Level 1	[360–320000] vs. [8.7e-07–0.037]	1.05	(0.90, 1.22)	0.5339
1994–1998	16	Exposure: Level 2 vs. Level 1	[23–160] vs. [0.12–23]	1.24	(1.08, 1.43)	0.0024
		Exposure: Level 3 vs. Level 1	[160–630] vs. [0.12–23]	1.27	(1.11, 1.45)	0.0005
		Exposure: Level 4 vs. Level 1	[630–1100] vs. [0.12–23]	1.26	(1.09, 1.46)	0.0019
		Exposure: Level 5 vs. Level 1	[1100–410,000] vs. [0.12–23]	0.98	(0.84, 1.14)	0.8043
1994–1998	24	Exposure: Level 2 vs. Level 1	[290–1100] vs. [2.2–290]	1.21	(1.06, 1.38)	0.0046
		Exposure: Level 3 vs. Level 1	[1100–1600] vs. [2.2–290]	1.26	(1.10, 1.44)	0.0007
		Exposure: Level 4 vs. Level 1	[1600–1900] vs. [2.2–290]	1.16	(1.00, 1.34)	0.0448
		Exposure: Level 5 vs. Level 1	[1900–440,000] vs. [2.2–290]	0.78	(0.67, 0.91)	0.0014
1994–2003	4	Exposure: Level 2 vs. Level 1	[9.8e-06–0.018] vs. [2e-13–9.8e-06]	0.96	(0.83, 1.11)	0.5606
		Exposure: Level 3 vs. Level 1	[0.018–19] vs. [2e-13–9.8e-06]	1.19	(1.04, 1.35)	0.0113
		Exposure: Level 4 vs. Level 1	[19–250] vs. [2e-13–9.8e-06]	1.20	(1.04, 1.39)	0.0116
		Exposure: Level 5 vs. Level 1	[250–230,000] vs. [2e-13–9.8e-06]	1.03	(0.89, 1.20)	0.6614
1994–2003	8	Exposure: Level 2 vs. Level 1	[0.17–11] vs. [2.9e-05–0.17]	1.00	(0.86, 1.16)	0.9911
		Exposure: Level 3 vs. Level 1	[11–270] vs. [2.9e-05–0.17]	1.19	(1.04, 1.36)	0.0105
		Exposure: Level 4 vs. Level 1	[270–1300] vs. [2.9e-05–0.17]	1.27	(1.10, 1.46)	0.0009
		Exposure: Level 5 vs. Level 1	[1300–370,000] vs. [2.9e-05–0.17]	1.06	(0.92, 1.23)	0.4385
1994–2003	16	Exposure: Level 2 vs. Level 1	[56–370] vs. [0.17–56]	1.23	(1.07, 1.42)	0.0043
		Exposure: Level 3 vs. Level 1	[370–1600] vs. [0.17–56]	1.17	(1.02, 1.35)	0.0271
		Exposure: Level 4 vs. Level 1	[1600–3900] vs. [0.17–56]	1.39	(1.20, 1.59)	<0.0001
		Exposure: Level 5 vs. Level 1	[3900–460,000] vs. [0.17–56]	1.05	(0.90, 1.22)	0.5115
1994–2003	24	Exposure: Level 2 vs. Level 1	[570–1800] vs. [3.9–570]	1.10	(0.96, 1.27)	0.1814
		Exposure: Level 3 vs. Level 1	[1800–4100] vs. [3.9–570]	1.24	(1.09, 1.42)	0.0013
		Exposure: Level 4 vs. Level 1	[4100–6200] vs. [3.9–570]	1.19	(1.03, 1.37)	0.0197
		Exposure: Level 5 vs. Level 1	[6200–500,000] vs. [3.9–570]	0.93	(0.80, 1.08)	0.3386

References

[1] R.M. Cohen, J.W. Mercer, DNAPL Site Evaluation, C.K. Smoley, Boca Raton, FL, 1993.
[2] M.T. Smith, Advances in understanding benzene health effects and susceptibility, Annu. Rev. Public Health 31 (2010) 133–148.
[3] N. Best, S. Cockings, J. Bennett, et al., Ecological regression analysis of environmental benzene exposure and childhood leukaemia: sensitivity to data inaccuracies, geographic scale and ecological bias, J. R. Stat. Soc. 164 (1) (2001) 155–174.
[4] IARC (International Agency for Research on Cancer), 2012. IARC monographs on the evaluation of carcinogenic risks to humans, volume 100f: benzene. Available: http://monographs.iarc.fr/ENG/Monographs/vol100F/mono100F-24.pdf (accessed 17.08.15).
[5] M.S. Linet, S. Yin, E.S. Gilbert, et al., A retrospective cohort study of cause-specific mortality and incidence of hematopoietic malignancies in Chinese benzene-exposed workers, Int. J. Cancer 137 (2015) 2184–2197.
[6] J. Vlaanderen, Q. Lan, H. Kromhout, et al., Occupational benzene exposure and the risk of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study quality dimensions, Environ. Health Perspect. 119 (2) (2011) 159–167.
[7] Facts about benzene, 2013. [Online]. Available: http://www.bt.cdc.gov/agent/benzene/basics/facts.asp (accessed 13.11.14).
[8] V. Bollati, A. Baccarelli, L. Hou, M. Bonzini, et al., Changes in DNA methylation patterns in subjects exposed to low-dose benzene, Cancer Res. 67 (3) (2007) 876–880.
[9] Q. Lan, L. Zhang, G. Li, et al., Hematotoxicity in workers exposed to low levels of benzene, Science 306 (5702) (2004) 1774–1776.
[10] L. Wallace, Major sources of benzene exposure, Environ. Health Perspect. 82 (1989) 165–169.
[11] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, CA: Cancer J. Clin. 65 (2015) 5–29.
[12] C.R. Flowers, J.O. Armitage, A decade of progress in lymphoma: advances and continuing challenges, Clin. Lymphoma Myeloma Leuk. 10 (6) (2010) 414–423.
[13] J.R. Cerhan, A. Kricker, O. Paltiel, et al., Medical history, lifestyle, family history, and occupational risk factors for diffuse large B-cell lymphoma: the InterLymph non-Hodgkin lymphoma subtypes project, J. Natl. Cancer Inst. Monogr. 2014 (2014) 15–25.
[14] M.S. Linet, C.M. Vajdic, L.M. Morton, et al., Medical history, lifestyle, family history, and occupational risk factors for follicular lymphoma: the InterLymph non-Hodgkin lymphoma subtypes project, J. Natl. Cancer Inst. Monogr. 2014 (2014) 26–40.
[15] L.M. Morton, S.L. Slager, J.R. Cerhan, et al., Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: the InterLymph non-Hodgkin lymphoma subtypes project, J. Natl. Cancer Inst. Monogr. 2014 (2014) 130–144.
[16] S.S. Wang, C.R. Flower, M.E. Kadin, et al., Medical history, lifestyle, family history, and occupational risk factors for peripheral T-cell lymphomas: the InterLymph non-Hodgkin lymphoma subtypes project, J. Natl. Cancer Inst. Monogr. 2014 (2014) 66–75.
[17] J.R. Cerhan, S.I. Berndt, J. Vijai, et al., Genome-wide association study identifies multiple susceptibility loci for diffuse large B-cell lymphoma, Nat. Genet. 2014 (2014) 1233–1238.
[18] N. Rothman, C.F. Skibola, S.S. Wang, et al., Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consort ium, Lancet Oncol. 7 (2006) 27–38.
[19] C.F. Skibola, P.M. Bracci, A. Nieters, et al., Tumor necrosis factor (TNF) and lymphotoxin-alpha (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph consortium, Am. J. Epidemiol. 171 (2010) 267–276.
[20] H.Y. Cao, P. Zou, H. Zhou, Genetic association of interleukin-10 promoter polymorphisms and susceptibility to diffuse large B-cell lymphoma: a meta-analysis, Gene 519 (2013) 288–294.
[21] W. Hu, B.A. Bassig, J. Xu, et al., Polymorphisms in pattern-recognition genes in the innate immunity system and risk of non-Hodgkin lymphoma, Environ. Mol. Mutagen. 54 (2013) 72–77.
[22] B.A. Bassig, J.R. Cerhan, W.Y. Au, et al., Genetic susceptibility to diffuse large B-cell lymphoma in a pooled study of three Eastern Asian populations, Eur. J. Haematol. 95 (5) (2015) 442–448.
[23] Y.Y. Sun, L. An, Y. Xie, et al., Methylenetetrahydrofolate reductase gene polymorphisms association with the risk of diffuse large B cell lymphoma: a meta-analysis, Tumour Biol. 34 (2013) 3587–3591.
[24] P. Cocco, A. t'Mannetje, D. Fadda, et al., Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study, Occup. Environ. Med. 67 (2010) 341–347.

[25] J. Vlaanderen, Q. Lan, N. Rothman, et al., Occupational benzene exposure and lymphoma risks, Environ. Health Perspect. 119 (11) (2011) a468–a469.
[26] A. Pronk, J. Nuckols, A.J. De Roos, et al., Residential proximity to industrial combustion facilities and risk of non-Hodgkin lymphoma: a case-control study, Environ. Health 12 (20) (2013).
[27] C. Bulka, L.J. Nastoupil, W. McClellan, et al., Residence proximity to benzene release sites is associated with increased incidence of non-Hodgkin lymphoma, Cancer 119 (18) (2013) 3309–3317.
[28] U.S. Census Bureau Introduction to Census 2000 Data Products, 2012.
[29] U.S. Census Bureau Census 2000 TIGER/Line Files: TIGER/Line Shapefiles and TIGER/Line Files, 2012.
[30] Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: Incidence-SEER 13 Registries Research Data, 2010.
[31] L.M. Morton, J.J. Turner, J.R. Cerhan, et al., Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consort ium (InterLymph), Blood 110 (2007) 695–708.
[32] Basics of TRI Reporting, United States Environmental Protection Agency, 29 September 2014. [Online]. Available: http://www2.epa.gov/toxics-release-inventory-tri-program/basics-tri-reporting (accessed 12.03.15).
[33] Toxics Release Inventory (TRI) Program, 2014. [Online]. Available: http://www2.epa.gov/toxics-release-inventory-tri-program (accessed 13.11.14).
[34] C.C. Robusto, The cosine-haversine formula, Am. Math. Mon. 64 (1) (1957) 38–40.
[35] D.C. Wheeler, L.A. Waller, W. Cozen, et al., Spatial-temporal analysis of non-Hodgkin lymphoma risk using multiple residential locations, Spat. Spatiotemporal Epidemiol. 3 (2) (2012) 163–171.
[36] A. Cameron, P. Trivedi, Regression Analysis of Count Data, Cambridge University Press, Cambridge, 1998.
[37] R: A language and environment for statistical computing, reference index verison 2.15.1, Vienna: R Core Development Team, 2014.
[38] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, Springer, New York, 2002.
[39] R. Bivand, N. Kewin-Koh, Maptools: tools for reading and handling spatial objects, 2014.
[40] J. Besag, Spatial interaction and the statistical analysis of lattice systems, J. R. Stat. Soc. Ser. B (Methodological) 2 (1974) 192–236.
[41] J. Besag, J. York, A. Mollie, Bayesian image restoration, with two applications in spatial statistics, Ann. Inst. Stat. Math. 43 (1991) 1–20.
[42] D. Clayton, J. Kaldor, Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, Biometrics 43 (1987) 671–681.